In [3]:
```python
import pandas as pd
```

In [4]:
```python
matches = pd.read_csv('matches.csv', index_col = 0)
matches.head()
```

Out[4]:

	date	time	comp	round	day	venue	result	gf	ga	opponent	...	match report	notes	sh	sot	dist	fk	pk	pkatt	se
1	2021-08-15	16:30	Premier League	Matchweek 1	Sun	Away	L	0.0	1.0	Tottenham	...	Match Report	NaN	18.0	4.0	16.9	1.0	0.0	0.0	
2	2021-08-21	15:00	Premier League	Matchweek 2	Sat	Home	W	5.0	0.0	Norwich City	...	Match Report	NaN	16.0	4.0	17.3	1.0	0.0	0.0	
3	2021-08-28	12:30	Premier League	Matchweek 3	Sat	Home	W	5.0	0.0	Arsenal	...	Match Report	NaN	25.0	10.0	14.3	0.0	0.0	0.0	
4	2021-09-11	15:00	Premier League	Matchweek 4	Sat	Away	W	1.0	0.0	Leicester City	...	Match Report	NaN	25.0	8.0	14.0	0.0	0.0	0.0	
6	2021-09-18	15:00	Premier League	Matchweek 5	Sat	Home	D	0.0	0.0	Southampton	...	Match Report	NaN	16.0	1.0	15.7	1.0	0.0	0.0	

5 rows × 27 columns

In [5]:
```python
matches.shape
```

Out[5]: (1389, 27)

In [6]:
```python
matches['team'].value_counts()
```

Out[6]:
```
Southampton                72
Brighton and Hove Albion   72
Manchester United          72
West Ham United            72
Newcastle United           72
Burnley                    71
Leeds United               71
Crystal Palace             71
Manchester City            71
Wolverhampton Wanderers    71
Tottenham Hotspur          71
Arsenal                    71
Leicester City             70
Chelsea                    70
Aston Villa                70
Everton                    70
Liverpool                  38
Fulham                     38
West Bromwich Albion       38
Sheffield United           38
Brentford                  34
Watford                    33
Norwich City               33
Name: team, dtype: int64
```

In [7]:
```python
1  matches[matches['team'] == 'Liverpool']
```

Out[7]:

	date	time	comp	round	day	venue	result	gf	ga	opponent	...	match report	notes	sh	sot	dist	fk	pk	pkatt
1	2020-09-12	17:30	Premier League	Matchweek 1	Sat	Home	W	4.0	3.0	Leeds United	...	Match Report	NaN	20.0	4.0	17.0	0.0	2.0	2.0
2	2020-09-20	16:30	Premier League	Matchweek 2	Sun	Away	W	2.0	0.0	Chelsea	...	Match Report	NaN	17.0	5.0	17.7	1.0	0.0	0.0
4	2020-09-28	20:00	Premier League	Matchweek 3	Mon	Home	W	3.0	1.0	Arsenal	...	Match Report	NaN	21.0	9.0	16.8	0.0	0.0	0.0
6	2020-10-04	19:15	Premier League	Matchweek 4	Sun	Away	L	2.0	7.0	Aston Villa	...	Match Report	NaN	14.0	8.0	15.8	1.0	0.0	0.0
7	2020-10-17	12:30	Premier League	Matchweek 5	Sat	Away	D	2.0	2.0	Everton	...	Match Report	NaN	22.0	8.0	15.0	1.0	0.0	0.0
9	2020-10-24	20:00	Premier League	Matchweek 6	Sat	Home	W	2.0	1.0	Sheffield Utd	...	Match Report	NaN	17.0	5.0	18.2	1.0	0.0	0.0
11	2020-10-31	17:30	Premier League	Matchweek 7	Sat	Home	W	2.0	1.0	West Ham	...	Match Report	NaN	8.0	2.0	18.6	1.0	1.0	1.0
13	2020-11-08	16:30	Premier League	Matchweek 8	Sun	Away	D	1.0	1.0	Manchester City	...	Match Report	NaN	9.0	2.0	21.5	0.0	1.0	1.0
14	2020-11-22	19:15	Premier League	Matchweek 9	Sun	Home	W	3.0	0.0	Leicester City	...	Match Report	NaN	24.0	12.0	11.9	0.0	0.0	0.0
16	2020-11-28	12:30	Premier League	Matchweek 10	Sat	Away	D	1.0	1.0	Brighton	...	Match Report	NaN	6.0	2.0	20.9	0.0	0.0	0.0
18	2020-12-06	19:15	Premier League	Matchweek 11	Sun	Home	W	4.0	0.0	Wolves	...	Match Report	NaN	11.0	6.0	16.6	1.0	0.0	0.0
20	2020-12-13	16:30	Premier League	Matchweek 12	Sun	Away	D	1.0	1.0	Fulham	...	Match Report	NaN	11.0	5.0	20.0	1.0	1.0	1.0
21	2020-12-16	20:00	Premier League	Matchweek 13	Wed	Home	W	2.0	1.0	Tottenham	...	Match Report	NaN	17.0	11.0	15.5	0.0	0.0	0.0
22	2020-12-19	12:30	Premier League	Matchweek 14	Sat	Away	W	7.0	0.0	Crystal Palace	...	Match Report	NaN	14.0	7.0	13.2	1.0	0.0	0.0
23	2020-12-27	16:30	Premier League	Matchweek 15	Sun	Home	D	1.0	1.0	West Brom	...	Match Report	NaN	17.0	2.0	17.8	2.0	0.0	0.0
24	2020-12-30	20:00	Premier League	Matchweek 16	Wed	Away	D	0.0	0.0	Newcastle Utd	...	Match Report	NaN	11.0	4.0	16.7	0.0	0.0	0.0

	date	time	comp	round	day	venue	result	gf	ga	opponent	...	match report	notes	sh	sot	dist	fk	pk	pkatt
25	2021-01-04	20:00	Premier League	Matchweek 17	Mon	Away	L	0.0	1.0	Southampton	...	Match Report	NaN	17.0	1.0	14.3	0.0	0.0	0.0
27	2021-01-17	16:30	Premier League	Matchweek 19	Sun	Home	D	0.0	0.0	Manchester Utd	...	Match Report	NaN	17.0	3.0	17.1	0.0	0.0	0.0
28	2021-01-21	20:00	Premier League	Matchweek 18	Thu	Home	L	0.0	1.0	Burnley	...	Match Report	NaN	27.0	6.0	17.3	0.0	0.0	0.0
30	2021-01-28	20:00	Premier League	Matchweek 20	Thu	Away	W	3.0	1.0	Tottenham	...	Match Report	NaN	14.0	7.0	14.7	0.0	0.0	0.0
31	2021-01-31	16:30	Premier League	Matchweek 21	Sun	Away	W	3.0	1.0	West Ham	...	Match Report	NaN	14.0	5.0	15.3	0.0	0.0	0.0
32	2021-02-03	20:15	Premier League	Matchweek 22	Wed	Home	L	0.0	1.0	Brighton	...	Match Report	NaN	11.0	0.0	19.9	0.0	0.0	0.0
33	2021-02-07	16:30	Premier League	Matchweek 23	Sun	Home	L	1.0	4.0	Manchester City	...	Match Report	NaN	7.0	2.0	17.9	1.0	1.0	1.0
34	2021-02-13	12:30	Premier League	Matchweek 24	Sat	Away	L	1.0	3.0	Leicester City	...	Match Report	NaN	15.0	4.0	15.4	2.0	0.0	0.0
36	2021-02-20	17:30	Premier League	Matchweek 25	Sat	Home	L	0.0	2.0	Everton	...	Match Report	NaN	16.0	6.0	15.9	1.0	0.0	0.0
37	2021-02-28	19:15	Premier League	Matchweek 26	Sun	Away	W	2.0	0.0	Sheffield Utd	...	Match Report	NaN	15.0	8.0	14.2	0.0	0.0	0.0
38	2021-03-04	20:15	Premier League	Matchweek 29	Thu	Home	L	0.0	1.0	Chelsea	...	Match Report	NaN	7.0	1.0	18.4	0.0	0.0	0.0
39	2021-03-07	14:00	Premier League	Matchweek 27	Sun	Home	L	0.0	1.0	Fulham	...	Match Report	NaN	16.0	3.0	17.0	1.0	0.0	0.0
41	2021-03-15	20:00	Premier League	Matchweek 28	Mon	Away	W	1.0	0.0	Wolves	...	Match Report	NaN	12.0	4.0	15.9	0.0	0.0	0.0
42	2021-04-03	20:00	Premier League	Matchweek 30	Sat	Away	W	3.0	0.0	Arsenal	...	Match Report	NaN	16.0	7.0	17.0	0.0	0.0	0.0
44	2021-04-10	15:00	Premier League	Matchweek 31	Sat	Home	W	2.0	1.0	Aston Villa	...	Match Report	NaN	23.0	8.0	16.7	1.0	0.0	0.0
46	2021-04-19	20:00	Premier League	Matchweek 32	Mon	Away	D	1.0	1.0	Leeds United	...	Match Report	NaN	17.0	7.0	15.7	0.0	0.0	0.0

	date	time	comp	round	day	venue	result	gf	ga	opponent	...	match report	notes	sh	sot	dist	fk	pk	pkatt
47	2021-04-24	12:30	Premier League	Matchweek 33	Sat	Home	D	1.0	1.0	Newcastle Utd	...	Match Report	NaN	21.0	9.0	17.0	1.0	0.0	0.0
48	2021-05-08	20:15	Premier League	Matchweek 35	Sat	Home	W	2.0	0.0	Southampton	...	Match Report	NaN	14.0	6.0	12.4	0.0	0.0	0.0
49	2021-05-13	20:15	Premier League	Matchweek 34	Thu	Away	W	4.0	2.0	Manchester Utd	...	Match Report	NaN	17.0	8.0	14.9	0.0	0.0	0.0
50	2021-05-16	16:30	Premier League	Matchweek 36	Sun	Away	W	2.0	1.0	West Brom	...	Match Report	NaN	26.0	6.0	16.9	1.0	0.0	0.0
51	2021-05-19	20:15	Premier League	Matchweek 37	Wed	Away	W	3.0	0.0	Burnley	...	Match Report	NaN	20.0	3.0	15.5	1.0	0.0	0.0
52	2021-05-23	16:00	Premier League	Matchweek 38	Sun	Home	W	2.0	0.0	Crystal Palace	...	Match Report	NaN	19.0	5.0	14.2	1.0	0.0	0.0

38 rows × 27 columns

In [8]:
```python
# matdches played in each match week

matches['round'].value_counts()
```

```
Out[8]:  Matchweek 1     39
         Matchweek 16    39
         Matchweek 34    39
         Matchweek 32    39
         Matchweek 31    39
         Matchweek 29    39
         Matchweek 28    39
         Matchweek 26    39
         Matchweek 25    39
         Matchweek 24    39
         Matchweek 23    39
         Matchweek 2     39
         Matchweek 19    39
         Matchweek 17    39
         Matchweek 20    39
         Matchweek 15    39
         Matchweek 5     39
         Matchweek 3     39
         Matchweek 13    39
         Matchweek 12    39
         Matchweek 4     39
         Matchweek 11    39
         Matchweek 10    39
         Matchweek 9     39
         Matchweek 8     39
         Matchweek 14    39
         Matchweek 7     39
         Matchweek 6     39
         Matchweek 30    37
         Matchweek 27    37
         Matchweek 22    37
         Matchweek 21    37
         Matchweek 18    37
         Matchweek 33    32
         Matchweek 35    20
         Matchweek 36    20
         Matchweek 37    20
         Matchweek 38    20
         Name: round, dtype: int64
```

Cleaning our data for machine learning

```
In [9]:    1  matches.dtypes
```

```
Out[9]:  date            object
         time            object
         comp            object
         round           object
         day             object
         venue           object
         result          object
         gf             float64
         ga             float64
         opponent        object
         xg             float64
         xga            float64
         poss           float64
         attendance     float64
         captain         object
         formation       object
         referee         object
         match report    object
         notes          float64
         sh             float64
         sot            float64
         dist           float64
         fk             float64
         pk             float64
         pkatt          float64
         season           int64
         team            object
         dtype: object
```
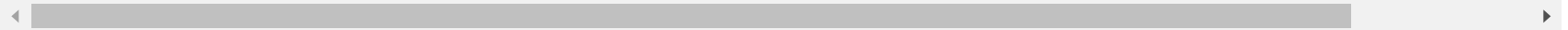
```
In [10]:   1  matches['date'] = pd.to_datetime(matches['date'])
```

In [11]:
```
1  matches.head()
```

Out[11]:

	date	time	comp	round	day	venue	result	gf	ga	opponent	...	match report	notes	sh	sot	dist	fk	pk	pkatt	se
1	2021-08-15	16:30	Premier League	Matchweek 1	Sun	Away	L	0.0	1.0	Tottenham	...	Match Report	NaN	18.0	4.0	16.9	1.0	0.0	0.0	
2	2021-08-21	15:00	Premier League	Matchweek 2	Sat	Home	W	5.0	0.0	Norwich City	...	Match Report	NaN	16.0	4.0	17.3	1.0	0.0	0.0	
3	2021-08-28	12:30	Premier League	Matchweek 3	Sat	Home	W	5.0	0.0	Arsenal	...	Match Report	NaN	25.0	10.0	14.3	0.0	0.0	0.0	
4	2021-09-11	15:00	Premier League	Matchweek 4	Sat	Away	W	1.0	0.0	Leicester City	...	Match Report	NaN	25.0	8.0	14.0	0.0	0.0	0.0	
6	2021-09-18	15:00	Premier League	Matchweek 5	Sat	Home	D	0.0	0.0	Southampton	...	Match Report	NaN	16.0	1.0	15.7	1.0	0.0	0.0	

5 rows × 27 columns

```
In [12]:    1  matches.dtypes
```

```
Out[12]:   date             datetime64[ns]
           time                     object
           comp                     object
           round                    object
           day                      object
           venue                    object
           result                   object
           gf                      float64
           ga                      float64
           opponent                 object
           xg                      float64
           xga                     float64
           poss                    float64
           attendance              float64
           captain                  object
           formation                object
           referee                  object
           match report            object
           notes                   float64
           sh                      float64
           sot                     float64
           dist                    float64
           fk                      float64
           pk                      float64
           pkatt                   float64
           season                    int64
           team                     object
           dtype: object
```

Creating predictors for machine learning

```
In [13]:    1  matches['venue_code'] = matches['venue'].astype('category').cat.codes
```

```
In [14]:   1  # venue code gavce 0 =away, 1 = home
           2  matches.head()
```

Out[14]:

	date	time	comp	round	day	venue	result	gf	ga	opponent	...	notes	sh	sot	dist	fk	pk	pkatt	season
1	2021-08-15	16:30	Premier League	Matchweek 1	Sun	Away	L	0.0	1.0	Tottenham	...	NaN	18.0	4.0	16.9	1.0	0.0	0.0	2022
2	2021-08-21	15:00	Premier League	Matchweek 2	Sat	Home	W	5.0	0.0	Norwich City	...	NaN	16.0	4.0	17.3	1.0	0.0	0.0	2022
3	2021-08-28	12:30	Premier League	Matchweek 3	Sat	Home	W	5.0	0.0	Arsenal	...	NaN	25.0	10.0	14.3	0.0	0.0	0.0	2022
4	2021-09-11	15:00	Premier League	Matchweek 4	Sat	Away	W	1.0	0.0	Leicester City	...	NaN	25.0	8.0	14.0	0.0	0.0	0.0	2022
6	2021-09-18	15:00	Premier League	Matchweek 5	Sat	Home	D	0.0	0.0	Southampton	...	NaN	16.0	1.0	15.7	1.0	0.0	0.0	2022

5 rows × 28 columns

```
In [15]:   1  matches['opp_code'] = matches['opponent'].astype('category').cat.codes
```

```
In [16]:   1  matches['hour'] = matches['time'].str.replace(':.+', '', regex=True).astype('int')
```

In [17]:
```
1 matches.head()
```

Out[17]:

	date	time	comp	round	day	venue	result	gf	ga	opponent	...	sot	dist	fk	pk	pkatt	season	team	ve
1	2021-08-15	16:30	Premier League	Matchweek 1	Sun	Away	L	0.0	1.0	Tottenham	...	4.0	16.9	1.0	0.0	0.0	2022	Manchester City	
2	2021-08-21	15:00	Premier League	Matchweek 2	Sat	Home	W	5.0	0.0	Norwich City	...	4.0	17.3	1.0	0.0	0.0	2022	Manchester City	
3	2021-08-28	12:30	Premier League	Matchweek 3	Sat	Home	W	5.0	0.0	Arsenal	...	10.0	14.3	0.0	0.0	0.0	2022	Manchester City	
4	2021-09-11	15:00	Premier League	Matchweek 4	Sat	Away	W	1.0	0.0	Leicester City	...	8.0	14.0	0.0	0.0	0.0	2022	Manchester City	
6	2021-09-18	15:00	Premier League	Matchweek 5	Sat	Home	D	0.0	0.0	Southampton	...	1.0	15.7	1.0	0.0	0.0	2022	Manchester City	

5 rows × 30 columns
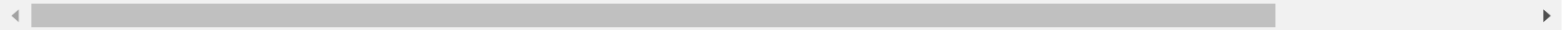
In [18]:
```
1 matches['day_code'] = matches['date'].dt.dayofweek
```

In [19]:
```
1 matches.head()
```

Out[19]:

	date	time	comp	round	day	venue	result	gf	ga	opponent	...	dist	fk	pk	pkatt	season	team	venue_
1	2021-08-15	16:30	Premier League	Matchweek 1	Sun	Away	L	0.0	1.0	Tottenham	...	16.9	1.0	0.0	0.0	2022	Manchester City	
2	2021-08-21	15:00	Premier League	Matchweek 2	Sat	Home	W	5.0	0.0	Norwich City	...	17.3	1.0	0.0	0.0	2022	Manchester City	
3	2021-08-28	12:30	Premier League	Matchweek 3	Sat	Home	W	5.0	0.0	Arsenal	...	14.3	0.0	0.0	0.0	2022	Manchester City	
4	2021-09-11	15:00	Premier League	Matchweek 4	Sat	Away	W	1.0	0.0	Leicester City	...	14.0	0.0	0.0	0.0	2022	Manchester City	
6	2021-09-18	15:00	Premier League	Matchweek 5	Sat	Home	D	0.0	0.0	Southampton	...	15.7	1.0	0.0	0.0	2022	Manchester City	

5 rows × 31 columns
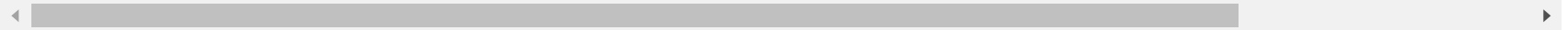
In [20]:
```
1 matches['target'] = (matches['result'] == 'W').astype('int')
```

In [21]:
```
1  matches.head()
```

Out[21]:

	date	time	comp	round	day	venue	result	gf	ga	opponent	...	fk	pk	pkatt	season	team	venue_code
1	2021-08-15	16:30	Premier League	Matchweek 1	Sun	Away	L	0.0	1.0	Tottenham	...	1.0	0.0	0.0	2022	Manchester City	0
2	2021-08-21	15:00	Premier League	Matchweek 2	Sat	Home	W	5.0	0.0	Norwich City	...	1.0	0.0	0.0	2022	Manchester City	1
3	2021-08-28	12:30	Premier League	Matchweek 3	Sat	Home	W	5.0	0.0	Arsenal	...	0.0	0.0	0.0	2022	Manchester City	1
4	2021-09-11	15:00	Premier League	Matchweek 4	Sat	Away	W	1.0	0.0	Leicester City	...	0.0	0.0	0.0	2022	Manchester City	0
6	2021-09-18	15:00	Premier League	Matchweek 5	Sat	Home	D	0.0	0.0	Southampton	...	1.0	0.0	0.0	2022	Manchester City	1

5 rows × 32 columns

Creating our initial machine learning model

In [22]:
```
1  from sklearn.ensemble import RandomForestClassifier
```

In [23]:
```
1  rf = RandomForestClassifier(n_estimators = 50, min_samples_split = 10, random_state = 1)
```

In [24]:
```
1  # this is a time series problem
```

In [25]:
```
1  train = matches[matches['date'] < '2022-01-01']
```

In [26]:
```
1  test = matches[matches['date'] > '2022-01-01']
```

In [27]:
```
1  predictors = ['venue_code', 'opp_code', 'hour', 'day_code']
```

```
In [28]:   1 rf.fit(train[predictors], train['target'])
```

Out[28]: RandomForestClassifier(min_samples_split=10, n_estimators=50, random_state=1)

```
In [29]:   1 preds = rf.predict(test[predictors])
```

```
In [30]:   1 from sklearn.metrics import accuracy_score
```

```
In [31]:   1 acc = accuracy_score(test['target'], preds)
```

```
In [32]:   1 acc
```

Out[32]: 0.6123188405797102

```
In [33]:   1 combined = pd.DataFrame(dict(actual=test['target'], prediction = preds))
```

```
In [34]:   1 pd.crosstab(index = combined['actual'], columns = combined['prediction'])
```

Out[34]:

prediction	0	1
actual		
0	141	31
1	76	28

```
In [ ]:   1 from sklearn.metrics import precision_score
```

```
In [36]:   1 precision_score(test['target'], preds)
```

Out[36]: 0.4745762711864407

Improving precision with rolling averages

the scores are low we need to improve this model

In [39]:
```python
1  grouped_matches =  matches.groupby('team')
```

In [41]:
```python
1  group = grouped_matches.get_group('Manchester City')   # we can check for other clubs
```

In [43]:
```python
1  group.head()
```

Out[43]:

	time	comp	round	day	venue	result	gf	ga	opponent	...	fk	pk	pkatt	season	team	venue_code	opp_code
	16:30	Premier League	Matchweek 1	Sun	Away	L	0.0	1.0	Tottenham	...	1.0	0.0	0.0	2022	Manchester City	0	18
	15:00	Premier League	Matchweek 2	Sat	Home	W	5.0	0.0	Norwich City	...	1.0	0.0	0.0	2022	Manchester City	1	15
	12:30	Premier League	Matchweek 3	Sat	Home	W	5.0	0.0	Arsenal	...	0.0	0.0	0.0	2022	Manchester City	1	0
	15:00	Premier League	Matchweek 4	Sat	Away	W	1.0	0.0	Leicester City	...	0.0	0.0	0.0	2022	Manchester City	0	10
	15:00	Premier League	Matchweek 5	Sat	Home	D	0.0	0.0	Southampton	...	1.0	0.0	0.0	2022	Manchester City	1	17

32 columns

In [44]:
```python
1  def rolling_averages(group, cols, new_cols):
2      group = group.sort_values('date')
3      rolling_stats = group[cols].rolling(3, closed = 'left').mean()
4      group[new_cols] = rolling_stats
5      group = group.dropna(subset = new_cols)
6      return group
```

In [45]:
```python
1  cols = ['gf', 'ga', 'sh', 'sot', 'dist', 'fk', 'pk', 'pkatt']
2  new_cols = [f'{c}_rolling' for c in cols]
```

In [46]:   1 new_cols

Out[46]: ['gf_rolling',
           'ga_rolling',
           'sh_rolling',
           'sot_rolling',
           'dist_rolling',
           'fk_rolling',
           'pk_rolling',
           'pkatt_rolling']

In [47]:
```
1  rolling_averages(group, cols, new_cols)
```

Out[47]:

	date	time	comp	round	day	venue	result	gf	ga	opponent	...	day_code	target	gf_rolling	ga_rolling	sh_rolling
5	2020-10-17	17:30	Premier League	Matchweek 5	Sat	Home	W	1.0	0.0	Arsenal	...	5	1	2.000000	2.333333	17.333333
7	2020-10-24	12:30	Premier League	Matchweek 6	Sat	Away	D	1.0	1.0	West Ham	...	5	0	1.333333	2.000000	17.333333
9	2020-10-31	12:30	Premier League	Matchweek 7	Sat	Away	W	1.0	0.0	Sheffield Utd	...	5	1	1.000000	0.666667	16.666667
11	2020-11-08	16:30	Premier League	Matchweek 8	Sun	Home	D	1.0	1.0	Liverpool	...	6	0	1.000000	0.333333	14.333333
12	2020-11-21	17:30	Premier League	Matchweek 9	Sat	Away	L	0.0	2.0	Tottenham	...	5	0	1.000000	0.666667	12.000000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
42	2022-03-14	20:00	Premier League	Matchweek 29	Mon	Away	D	0.0	0.0	Crystal Palace	...	0	0	2.333333	1.333333	19.000000
44	2022-04-02	15:00	Premier League	Matchweek 31	Sat	Away	W	2.0	0.0	Burnley	...	5	1	1.666667	0.333333	18.333333
46	2022-04-10	16:30	Premier League	Matchweek 32	Sun	Home	D	2.0	2.0	Liverpool	...	6	0	2.000000	0.333333	20.000000
49	2022-04-20	20:00	Premier League	Matchweek 30	Wed	Home	W	3.0	0.0	Brighton	...	2	1	1.333333	0.666667	15.666667
50	2022-04-23	15:00	Premier League	Matchweek 34	Sat	Home	W	5.0	1.0	Watford	...	5	1	2.333333	0.666667	15.333333

68 rows × 40 columns

In [48]:
```
1  matches_rolling = matches.groupby('team').apply(lambda x: rolling_averages(x, cols, new_cols))
```

In [49]:
```
1  matches_rolling
```

Out[49]:

team		date	time	comp	round	day	venue	result	gf	ga	opponent	...	day_code	target	gf_rolling	ga_ro
Arsenal	**6**	2020-10-04	14:00	Premier League	Matchweek 4	Sun	Home	W	2.0	1.0	Sheffield Utd	...	6	1	2.000000	1.333
	7	2020-10-17	17:30	Premier League	Matchweek 5	Sat	Away	L	0.0	1.0	Manchester City	...	5	0	1.666667	1.666
	9	2020-10-25	19:15	Premier League	Matchweek 6	Sun	Home	L	0.0	1.0	Leicester City	...	6	0	1.000000	1.666
	11	2020-11-01	16:30	Premier League	Matchweek 7	Sun	Away	W	1.0	0.0	Manchester Utd	...	6	1	0.666667	1.000
	13	2020-11-08	19:15	Premier League	Matchweek 8	Sun	Home	L	0.0	3.0	Aston Villa	...	6	0	0.333333	0.666
...	**...**	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Wolverhampton Wanderers	**32**	2022-03-13	14:00	Premier League	Matchweek 29	Sun	Away	W	1.0	0.0	Everton	...	6	1	1.333333	1.000
	33	2022-03-18	20:00	Premier League	Matchweek 30	Fri	Home	L	2.0	3.0	Leeds United	...	4	0	1.666667	0.666
	34	2022-04-02	15:00	Premier League	Matchweek 31	Sat	Home	W	2.0	1.0	Aston Villa	...	5	1	2.333333	1.000
	35	2022-04-08	20:00	Premier League	Matchweek 32	Fri	Away	L	0.0	1.0	Newcastle Utd	...	4	0	1.666667	1.333
	36	2022-04-24	14:00	Premier League	Matchweek 34	Sun	Away	L	0.0	1.0	Burnley	...	6	0	1.333333	1.666

1317 rows × 40 columns

In [50]:
```
1  matches_rolling = matches_rolling.droplevel('team')
```

In [51]:
```
1  matches_rolling
```

Out[51]:

	date	time	comp	round	day	venue	result	gf	ga	opponent	...	day_code	target	gf_rolling	ga_rolling	sh_rolling
6	2020-10-04	14:00	Premier League	Matchweek 4	Sun	Home	W	2.0	1.0	Sheffield Utd	...	6	1	2.000000	1.333333	7.666667
7	2020-10-17	17:30	Premier League	Matchweek 5	Sat	Away	L	0.0	1.0	Manchester City	...	5	0	1.666667	1.666667	5.333333
9	2020-10-25	19:15	Premier League	Matchweek 6	Sun	Home	L	0.0	1.0	Leicester City	...	6	0	1.000000	1.666667	7.000000
11	2020-11-01	16:30	Premier League	Matchweek 7	Sun	Away	W	1.0	0.0	Manchester Utd	...	6	1	0.666667	1.000000	9.666667
13	2020-11-08	19:15	Premier League	Matchweek 8	Sun	Home	L	0.0	3.0	Aston Villa	...	6	0	0.333333	0.666667	9.666667
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
32	2022-03-13	14:00	Premier League	Matchweek 29	Sun	Away	W	1.0	0.0	Everton	...	6	1	1.333333	1.000000	12.333333
33	2022-03-18	20:00	Premier League	Matchweek 30	Fri	Home	L	2.0	3.0	Leeds United	...	4	0	1.666667	0.666667	12.333333
34	2022-04-02	15:00	Premier League	Matchweek 31	Sat	Home	W	2.0	1.0	Aston Villa	...	5	1	2.333333	1.000000	13.000000
35	2022-04-08	20:00	Premier League	Matchweek 32	Fri	Away	L	0.0	1.0	Newcastle Utd	...	4	0	1.666667	1.333333	13.000000
36	2022-04-24	14:00	Premier League	Matchweek 34	Sun	Away	L	0.0	1.0	Burnley	...	6	0	1.333333	1.666667	10.000000

1317 rows × 40 columns

In [52]:
```
1  matches_rolling.index = range(matches_rolling.shape[0])
```

In [53]: 1 matches_rolling

Out[53]:

	date	time	comp	round	day	venue	result	gf	ga	opponent	...	day_code	target	gf_rolling	ga_rolling	sh_rollin
0	2020-10-04	14:00	Premier League	Matchweek 4	Sun	Home	W	2.0	1.0	Sheffield Utd	...	6	1	2.000000	1.333333	7.66666
1	2020-10-17	17:30	Premier League	Matchweek 5	Sat	Away	L	0.0	1.0	Manchester City	...	5	0	1.666667	1.666667	5.33333
2	2020-10-25	19:15	Premier League	Matchweek 6	Sun	Home	L	0.0	1.0	Leicester City	...	6	0	1.000000	1.666667	7.00000
3	2020-11-01	16:30	Premier League	Matchweek 7	Sun	Away	W	1.0	0.0	Manchester Utd	...	6	1	0.666667	1.000000	9.66666
4	2020-11-08	19:15	Premier League	Matchweek 8	Sun	Home	L	0.0	3.0	Aston Villa	...	6	0	0.333333	0.666667	9.66666
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	.
1312	2022-03-13	14:00	Premier League	Matchweek 29	Sun	Away	W	1.0	0.0	Everton	...	6	1	1.333333	1.000000	12.33333
1313	2022-03-18	20:00	Premier League	Matchweek 30	Fri	Home	L	2.0	3.0	Leeds United	...	4	0	1.666667	0.666667	12.33333
1314	2022-04-02	15:00	Premier League	Matchweek 31	Sat	Home	W	2.0	1.0	Aston Villa	...	5	1	2.333333	1.000000	13.00000
1315	2022-04-08	20:00	Premier League	Matchweek 32	Fri	Away	L	0.0	1.0	Newcastle Utd	...	4	0	1.666667	1.333333	13.00000
1316	2022-04-24	14:00	Premier League	Matchweek 34	Sun	Away	L	0.0	1.0	Burnley	...	6	0	1.333333	1.666667	10.00000

1317 rows × 40 columns

Retraining our machine learning model

In [54]:
```python
def make_predictions(data, predictors):
    train = data[data['date'] < '2022-01-01']
    test = data[data['date'] > '2022-01-01']
    rf.fit(train[predictors], train['target'])
    preds = rf.predict(test[predictors])
    combined = pd.DataFrame(dict(actual = test['target'], predicted = preds), index = test.index)
    precision = precision_score(test['target'], preds)
    return combined, precision
```

In [56]:
```python
combined, precision = make_predictions(matches_rolling, predictors + new_cols)
```

In [57]:
```python
precision
```

Out[57]: 0.625

In [58]:
```python
combined
```

Out[58]:

	actual	predicted
55	0	0
56	1	0
57	1	0
58	1	1
59	1	1
...	...	...
1312	1	0
1313	0	0
1314	1	0
1315	0	0
1316	0	0

276 rows × 2 columns

In [60]:
```
1 combined = combined.merge(matches_rolling[['date', 'team', 'opponent', 'result']], left_index = True,
```

In [61]:
```
1 combined
```

Out[61]:

	actual	predicted	date	team	opponent	result
55	0	0	2022-01-23	Arsenal	Burnley	D
56	1	0	2022-02-10	Arsenal	Wolves	W
57	1	0	2022-02-19	Arsenal	Brentford	W
58	1	1	2022-02-24	Arsenal	Wolves	W
59	1	1	2022-03-06	Arsenal	Watford	W
...	...	...	...	...	...	...
1312	1	0	2022-03-13	Wolverhampton Wanderers	Everton	W
1313	0	0	2022-03-18	Wolverhampton Wanderers	Leeds United	L
1314	1	0	2022-04-02	Wolverhampton Wanderers	Aston Villa	W
1315	0	0	2022-04-08	Wolverhampton Wanderers	Newcastle Utd	L
1316	0	0	2022-04-24	Wolverhampton Wanderers	Burnley	L

276 rows × 6 columns

Combining home and away prediction

```python
In [62]:   1  class MissingDict(dict):
           2      __missing__ = lambda self, key: key
           3
           4  map_values = {
           5      'Brighton and Hove Albion': 'Brighton',
           6      'Manchester United': 'Manchester Utd',
           7      'Newcastle United': 'Newcastle Utd',
           8      'Tottenham Hotspur': 'Tottenham',
           9      'West Ham United': 'West Ham',
          10      'Wolverhampton Wanderers': 'Wolves'
          11  }
          12  mapping = MissingDict(**map_values)
```

```python
In [68]:   1  mapping['West Ham United'], mapping['Arsenal']
```

```
Out[68]:  ('West Ham', 'Arsenal')
```

```python
In [69]:   1  combined['new_team'] = combined['team'].map(mapping)
```

In [70]:
```
1 combined
```

Out[70]:

	actual	predicted	date	team	opponent	result	new_team
55	0	0	2022-01-23	Arsenal	Burnley	D	Arsenal
56	1	0	2022-02-10	Arsenal	Wolves	W	Arsenal
57	1	0	2022-02-19	Arsenal	Brentford	W	Arsenal
58	1	1	2022-02-24	Arsenal	Wolves	W	Arsenal
59	1	1	2022-03-06	Arsenal	Watford	W	Arsenal
...	...	...	...	...	...	...	...
1312	1	0	2022-03-13	Wolverhampton Wanderers	Everton	W	Wolves
1313	0	0	2022-03-18	Wolverhampton Wanderers	Leeds United	L	Wolves
1314	1	0	2022-04-02	Wolverhampton Wanderers	Aston Villa	W	Wolves
1315	0	0	2022-04-08	Wolverhampton Wanderers	Newcastle Utd	L	Wolves
1316	0	0	2022-04-24	Wolverhampton Wanderers	Burnley	L	Wolves

276 rows × 7 columns

In [72]:
```
1 merged = combined.merge(combined, left_on = ['date','new_team'], right_on = ['date', 'opponent'])
```

In [73]: | 1 | merged |

Out[73]:

	actual_x	predicted_x	date	team_x	opponent_x	result_x	new_team_x	actual_y	predicted_y	team_y	opponent_
0	0	0	2022-01-23	Arsenal	Burnley	D	Arsenal	0	0	Burnley	Arsenal
1	1	0	2022-02-10	Arsenal	Wolves	W	Arsenal	0	0	Wolverhampton Wanderers	Arsenal
2	1	0	2022-02-19	Arsenal	Brentford	W	Arsenal	0	0	Brentford	Arsenal
3	1	1	2022-02-24	Arsenal	Wolves	W	Arsenal	0	0	Wolverhampton Wanderers	Arsenal
4	1	1	2022-03-06	Arsenal	Watford	W	Arsenal	0	0	Watford	Arsenal
...	...	...	...	...	...	...	...	...	...	...	
257	1	0	2022-03-13	Wolverhampton Wanderers	Everton	W	Wolves	0	0	Everton	Wolve
258	0	0	2022-03-18	Wolverhampton Wanderers	Leeds United	L	Wolves	1	0	Leeds United	Wolve
259	1	0	2022-04-02	Wolverhampton Wanderers	Aston Villa	W	Wolves	0	0	Aston Villa	Wolve
260	0	0	2022-04-08	Wolverhampton Wanderers	Newcastle Utd	L	Wolves	1	0	Newcastle United	Wolve
261	0	0	2022-04-24	Wolverhampton Wanderers	Burnley	L	Wolves	1	0	Burnley	Wolve

262 rows × 13 columns

In [74]: | 1 | merged[(merged['predicted_x'] == 1) & (merged['predicted_y'] == 0)]['actual_x'].value_counts() |

Out[74]:
```
1    27
0    13
Name: actual_x, dtype: int64
```

In [76]:
```
1  27/40
```

Out[76]: 0.675

In [77]:
```
1  matches.columns
```

Out[77]: Index(['date', 'time', 'comp', 'round', 'day', 'venue', 'result', 'gf', 'ga',
              'opponent', 'xg', 'xga', 'poss', 'attendance', 'captain', 'formation',
              'referee', 'match report', 'notes', 'sh', 'sot', 'dist', 'fk', 'pk',
              'pkatt', 'season', 'team', 'venue_code', 'opp_code', 'hour', 'day_code',
              'target'],
             dtype='object')

In [ ]:
```
1  # we only used some columns for this machine learning
```